

MHSNet: A Multi-Scale Hidden State Interaction Network for Fault Diagnosis of Rotating Machinery

Fan Zhang, Zhaoqi Li, Yao Cheng*, Yan Yang, Yufei Han*, Cai Yi, Tianrui Li, Jingke Yan, Likai Dong, and Weihua Zhang

Abstract: The rotating machinery system consists of several key components such as bearings and gears. The operating condition of the bearings directly affects equipment safety and production efficiency. However, traditional bearing fault diagnosis methods face challenges in complex operating conditions, including insufficient local feature extraction, severe noise interference, and difficulty in integrating global information due to the heterogeneity of multi-sensor data. To address these issues, this paper proposes a multi-sensor and multi-task fault diagnosis method based on the Multi-Scale Hidden State Interaction Network (MHSNet). In terms of feature extraction, MHSNet integrates deep separable convolutions with hidden state-space models. By introducing multi-scale convolution units, it captures local details under different receptive fields. Additionally, the selective hidden state modeling mechanism of the Mamba module overcomes the limitations of conventional convolution networks' local receptive fields, enabling the modeling of periodic impulses and long-range dependencies in signals. In the data fusion layer, a dynamic state space fusion module is designed to achieve parameterized interaction and adaptive alignment of multi-sensor data within the hidden state space, effectively alleviating the distribution differences and redundancy issues between multi-source information. Through the collaborative extraction of complementary features between tasks, the model further enhances robustness and discriminative accuracy under conditions of data imbalance and noise interference. Extensive experiments conducted on real bearing data and multi-condition testing platforms demonstrate that MHSNet consistently achieves high diagnostic accuracy and condition classification performance. It outperforms traditional unimodal and heterogeneous multi-sensor signal-based diagnostic networks, highlighting its significant advantages in multi-sensor collaborative representation, global and local feature fusion, and noise suppression.

Key words: bearing fault detection; deep learning; multi-sensor; multi-task; mamba

1 Introduction

With the advent of Industry 4.0, rotating machinery has become pivotal in modern industrial systems. Damage to key components such as bearings and gears can lead to equipment failure, production interruptions, and safety hazards, substantially increasing production and maintenance costs and threatening human health and safety [1]. However,

harsh operating environments often cause transient impulsive signals to be submerged in large amounts of non-fault-related information. Consequently, timely and accurate condition monitoring and fault diagnosis of critical components in rotating machinery are essential for maintaining and optimizing industrial processes [2].

Over the past decades, traditional vibration-analysis-based fault detection methods have been widely employed for bearing fault detection [3]. Nevertheless, these methods place stringent demands on the design

of morphological operators; when the design is suboptimal, it is often difficult to extract impulsive signatures buried in high-frequency noise, which markedly degrades detection accuracy. With the rapid progress of machine learning, ML-based diagnostic approaches have gradually become a major direction in rotating machinery diagnostics [4]. Classical techniques such as support vector machines (SVMs) [5], the discrete wavelet transform (DWT) [6], and the fast Fourier transform (FFT) [7] have been widely applied. However, these approaches typically rely on expert prior knowledge and handcrafted feature extraction, limiting their ability to automatically discover high-level representations and, in turn, constraining further improvements in diagnostic accuracy [8].

In recent years, deep learning has achieved notable success across computer science and technology [9, 10], mechanical engineering [11], electrical engineering [12, 13], and materials science [14]. Owing to its powerful capabilities in representation learning and pattern recognition, deep-learning-based fault diagnosis methods [15] have been increasingly adopted. For example, Han et al. [16] proposed a dual-domain adversarial residual encoder–decoder network employing a cascaded

architecture of adversarial residual encoder–decoders for denoising and feature extraction. To address the limited generalization of Transformers in fault diagnosis, Xiao et al. [17] introduced a Bayesian variational Transformer–based diagnostic approach and achieved competitive performance. Nevertheless, for complete and complex industrial systems, single-sensor approaches are insufficient to comprehensively assess system operating conditions and health status, and their performance in such scenarios remains suboptimal.

Compared with single-sensor data, multi-sensor signals combine features from multiple modalities and thus contain rich and complementary fault information [18]. Consequently, multi-sensor–based diagnostic methods have attracted growing interest [19]. However, many fusion strategies rely on early or late fusion and often overlook distributional discrepancies and complex interactions among heterogeneous signals [20]. With the rise of intermediate-layer fusion strategies in computer vision [21], researchers have found that mid-level fusion can exploit cross-modal correlations among heterogeneous sensors to realize multi-level cross-modal integration. Xu et al. [22], considering intrinsic correlations and distribution shifts across signals, proposed a co-fusion convolutional neural network that, for the first time, explored intrinsic associations via intermediate-layer fusion and integrated cross-modal features. Xue et al. [23] further developed a method that fuses multi-transform-domain and multi-source data to extract key features from multiple domains and fusion layers. While these CNN-based approaches leverage sophisticated attention mechanisms, they are fundamentally limited by the finite receptive field of convolutional units and struggle to capture long-range dependencies.

By contrast, Transformers excel at modeling global dependencies in long-sequence tasks [24], but the quadratic complexity of self-attention hampers efficiency for long sequences and high sampling rates. Recently, state-space models (SSMs) [25], epitomized by Mamba, have emerged as an effective alternative for long-sequence modeling, achieving long-range dependency representation with linear complexity. Prior studies [26] indicate that Mamba offers advantages over conventional CNNs/Transformers in effective receptive field and scalability by rephrasing attention as state evolution that scales linearly with sequence length, thereby markedly reducing the computation cost for long sequences. In multi-sensor

-
- Fan Zhang is with the School of Design, and Sichuan Social Sciences Key Laboratory of Intelligent Design for Complex Transportation Systems, Southwest Jiaotong University, Chengdu 611756, China. E-mail: fan.zhang@swjtu.edu.cn
 - Zhaoqi Li, Yan Yang and Tianrui Li are with the School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China. E-mail: lizhaoqi@my.swjtu.edu.cn; yyang@swjtu.edu.cn; trli@swjtu.edu.cn, h1456214699@gmail.com
 - Yufei Han is with the Information System Integration Center, China Mobile Communications Group Shandong Company Limited, Jinan, China. E-mail: h1456214699@gmail.com
 - Yao Cheng, Cai Yi, Jingke Yan and Weihua Zhang are with the State Key Laboratory of Rail Transit Vehicle System, Southwest Jiaotong University, Chengdu 610000, China. E-mail: chengyao2020@swjtu.edu.cn; yicai@swjtu.edu.cn; yanjingke@my.swjtu.edu.cn; tpl@swjtu.edu.cn
 - Likai Dong is with the School of Information Science and Engineering, University of Jinan, Jinan 250022, China. E-mail: ise_donglk@ujn.edu.cn

* Correspondence to: Yao Cheng(chengyao2020@swjtu.edu.cn), Yufei Han(h1456214699@gmail.com).

Manuscript received: 2025-05-19; revised: 2025-08-17; accepted: 2024-08-28

scenarios, Mamba enables mid-level cross-modal interaction at the hidden-state level: on the one hand aligning and enhancing local salient cues such as impulsive events, and on the other characterizing intrinsic inter-modal correlations from the perspectives of global periodicity and long-range dependencies, thus overcoming the limitations of traditional attention weighting that often fails to jointly model “local-global” relations [27]. Meanwhile, the stable local inductive bias of CNNs provides reliable priors for Mamba and suppresses spurious activations during long-range modeling. Based on this insight, designing hidden-state interactions between CNNs and Mamba unifies fine-grained local features and global long-range dependencies within a single modeling framework.

In practical industrial systems, it is often necessary to monitor both fault conditions and operating states simultaneously; multitask learning (MTL) provides a feasible solution [28]. For instance, Cui et al. [29] extracted and fused features from raw vibration and current signals using CNNs and constructed a multitask diagnostic network comprising classification and deep metric-learning tasks. Addressing the difficulty of fully extracting complementary fault features from multi-sensor data, Wang et al. [30] introduced a feature-level attention module to adaptively learn task-specific representations from globally shared features, enabling joint modeling of bearing fault diagnosis and feature-source discrimination. However, existing multitask diagnostic methods typically route fused features directly to task-specific branches without adequately accounting for differences in sensor importance across tasks. In practice, this omission leads to two issues: first, when fault signals are heavily contaminated by noise, models struggle to suppress the impact of noise on feature extraction and decision-making; second, the amount of data available under fault conditions is often far smaller than under normal operation. The resulting class imbalance yields insufficient training samples and skewed label distributions, increasing the risk of overfitting or classifier bias and ultimately degrading fault-recognition accuracy. Therefore, efficiently fusing multi-source data becomes a key challenge. To this end, a Mamba-based hidden state-space fusion approach can align and interact across sensor modalities at the hidden-state level, model long-term dependencies via a learnable state-transition mechanism, and adaptively capture inter-sensor relationships through hidden-state correlations, thereby enhancing the robustness and

generalization of multitask diagnosis.

Motivated by the above, this paper proposes a Multi-scale Hidden State Interaction Network (MHSNet) that performs dynamic fusion and interaction between global contextual information and local salient cues across multiple sensors, while enabling task-aware selection of features from heterogeneous multi-sensor signals.

The main contributions are as follows:

- (1) We introduce a diagnostic network for heterogeneous multi-sensor signals that models discrepancies in local salient information, captures intrinsic inter-signal correlations from periodic variation and global dependencies, and explicitly accounts for the differing importance of sensors across tasks, thereby strengthening feature selection within each task branch.
- (2) By combining the strengths of CNNs and Mamba, we design a Mamba-based Multi-Scale Module (MMS) that fuses global, periodic dependencies with multi-scale local feature cues across sensors to facilitate cross-modal interaction.
- (3) We develop a Mamba-based Dynamic State Space Fusion (DSSF) module with an adaptive selection mechanism that fully accounts for potential heterogeneity among sensors, achieving effective cross-source alignment and thorough exploitation of global associations.

2 Related Work

2.1 Traditional Signal Processing and Machine Learning Methods

In the field of rotating machinery fault diagnosis, traditional approaches primarily rely on time-domain, frequency-domain, and time-frequency analysis techniques [3]. Methods such as blind deconvolution, spectral coherence, sparsity metrics, and morphological filtering can extract periodic and modulation features from vibration signals, but exhibit limited robustness in the presence of strong noise and non-stationary signals [31].

To enhance transient pulse feature extraction, Li et al. [32] proposed a morphological gradient convolution combined with a cross-correlation operator, which leverages the periodicity-enhancing properties of convolution and correlation to highlight transient signals. However, under non-Gaussian noise and

complex periodic backgrounds, fault-related features remain vulnerable to interference [33]. Chen et al. [34] addressed the dependency of structural-element parameters on manual experience and the lack of adaptivity by introducing an adaptive transient feature extractor based on cubic spline interpolation, which automatically adjusts the SE parameters according to the vibration signal, thereby improving feature matching capability. Nevertheless, this method still lacks sufficient robustness under complex operating conditions and heavy noise. Furthermore, multi-scale morphological filtering [35] and optimal-scale morphological filtering [36] improve adaptability by introducing structural elements of varying lengths and shapes, but they generally suffer from high computational cost and poor real-time performance.

Regarding the cyclostationary characteristics of fault signals, Borghesani et al. [37] proposed a hybrid model that combines cyclostationary and pseudo-cyclostationary signal modeling to enhance feature extraction for rolling bearings. Antoni et al. [38] and Randall et al. [39] developed spectral correlation and spectral coherence analysis methods for joint visualization of modulation and carrier frequencies, advancing the development of cyclostationary-based analytical tools. However, under low signal-to-noise ratio and complex frequency bands, characteristic frequencies are easily submerged by noise. To address this, Mauricio et al. [40] and Schmidt et al. [41] introduced weighted envelope spectra (e.g., CIES, WCES) and adaptive band-selection algorithms to improve the effectiveness of feature-band extraction, but these techniques still depend on experiential selection of weight parameters and thresholds, and exhibit limited automation [42]. Building on these feature-extraction methods, Hosseini et al. [5] and Arevalo et al. [6] applied machine learning classifiers such as support vector machines (SVM), k-nearest neighbors (KNN), decision trees, and random forests to automatically classify handcrafted features, thereby improving diagnostic efficiency. However, the performance of these methods is highly contingent on the efficacy of feature engineering, and they struggle to deliver stable diagnostic results in high-dimensional, high-noise, or complex operating scenarios.

In summary, although traditional signal-processing and machine-learning methods have achieved incremental progress in rotating machinery fault diagnosis, they commonly suffer from manual feature-

extraction dependency, complex parameter tuning, poor noise robustness, and low levels of automation, rendering them ill-suited to the variable conditions and strong-noise environments of modern industrial applications. These limitations have driven the emergence of intelligent approaches based on deep learning and multi-sensor fusion.

2.2 Multi-Sensor Fusion Based on Deep Learning

Deep learning models[43]—including CNNs, RNNs and Transformers—have significantly improved automatic feature extraction from one-dimensional vibration and acoustic signals, reducing reliance on handcrafted priors [44]. Architectures such as multi-scale feature networks, deep residual shrinkage for high-noise scenarios, and lightweight designs for small-sample settings demonstrate robustness gains, yet they still demand large annotated datasets and can suffer under severe noise or distribution shifts [45].

To overcome the limitations of single-sensor information, multi-sensor data fusion has gradually become a research hotspot. Multi-sensor fusion not only combines rich multi-modal features to enhance recognition of complex faults, but also improves system robustness and reliability [46]. For instance, Liu et al.[47] proposed an early-fusion CNN for multi-channel vibration signals, achieving improved diagnostic accuracy under complex operating conditions; Xiao et al.[48] designed a three-stage feature fusion strategy to fully integrate vibration and torque signal information; He et al.[49] employed transfer learning and adversarial fusion to achieve highly robust fault detection across domains. Yet most fusion methods focus on a specific level (e.g., data-level, feature-level, or decision-level) and thus fail to fully capture global contextual dependencies and complex interactions among modalities [20]. In designing deep-fusion mechanisms, some researchers have used sparse autoencoders, deep belief networks, and one-dimensional convolutional LSTM models for multi-source data fusion to enable end-to-end feature extraction and diagnosis. For example, Chen et al.[50] employed sparse autoencoders and deep belief networks for feature fusion; Wang et al.[51] optimized CNN structures via a bottleneck layer to enhance diagnostic performance in multi-vibration-sensor fusion; Hao et al.[52] integrated multi-sensor signals using LSTM, effectively improving recognition rates in complex scenarios. However, these methods

still require improvements in model generalization and long-term dependency modeling to handle nonlinear, non-stationary, and high-noise interference signals in real industrial environments.

In summary, deep learning methods and multi-sensor fusion strategies have significantly advanced the automation and accuracy of rotating machinery fault diagnosis. Although existing approaches have achieved breakthroughs in automatic feature extraction, multi-source data utilization, and system robustness, key challenges remain in achieving more efficient, robust, and generalizable multi-modal information fusion under small-sample, noise-interference, and multi-condition distribution-shift scenarios, as well as in better modeling of long-term global dependencies and complex inter-modal interactions.

2.3 Multi-Task Learning and Existing Challenges

As industrial application scenarios become increasingly complex, single-task models struggle to meet the precise discrimination requirements of multi-dimensional equipment health states. MTL provides an effective theoretical foundation for collaboratively optimizing multiple objectives such as fault-type identification and operating-condition classification [28]. Current MTL frameworks typically adopt shared low-level feature representations or introduce task-specific branches to enhance the model's discriminative power. Guo et al.[53] constructed a multi-task CNN to simultaneously identify rolling bearing fault types and locations. Xie et al.[54] applied an MTL strategy to jointly recognize fault types and severities under small-sample constraints. Hasan et al.[55] fused raw signals, spectra, and envelope spectra into two-dimensional images and utilized an MTL framework for speed recognition and bearing fault diagnosis. Niu et al.[56] improved joint discrimination of fault types and locations through a multi-task approach.

Despite these advances, several challenges remain. First, accurately modeling the correlations and heterogeneities among different modalities and tasks is difficult; in practice, models are susceptible to noise interference and data imbalance, leading to degraded generalization performance [28]. Most existing multi-task fusion strategies focus solely on feature concatenation, neglecting dynamic interactions between global and local features and selective information transfer, with limited ability to model

long-term dependencies. Meanwhile, current research mostly concentrates on multi-health information recognition for a single component or a single fault type; studies on jointly extracting shared and private features among different signal components in compound-fault scenarios, as well as on suppressing task interference, remain scarce.

In summary, designing a highly robust intelligent diagnostic network that combines multi-modal dynamic fusion with multi-task discrimination, fully exploits the effective associations between tasks and signal sources under complex working conditions, and enhances model generalization under noise and data-imbalance conditions, is a critical scientific challenge in the field of intelligent fault diagnosis.

3 Methods

3.1 Framework of MHSNet

To effectively extract the time-domain features of multi-sensor signals and achieve intelligent fault diagnosis, this paper proposes a multi-sensor and multi-task fault diagnosis method based on the Multi-Scale Hidden State Interaction Network (MHSNet). Fig. 1 shows the overall framework of the proposed method. The framework consists of four core components: multi-sensor data acquisition from the test bench, Mamba-based Multi-Scale Module (MMS), Dynamic State Space Fusion (DSSF) Module, and multi-task classification.

First, given the multi-sensor time-series data x^1, x^2 , each data is preprocessed and sliced into fixed-length segments. The multi-scale feature encoding u^i of the single-sensor signal is obtained through the designed MMS, with the mapping relationship as shown in Eq. (1):

$$u^i = \text{MMS}(x^i; \Theta_{\text{MMS}}^i), i \in \{1, 2\} \quad (1)$$

where Θ_{MMS}^i denotes the parameters that the MMS branch network for the i -th sensor must learn. The MMS module is composed primarily of multi-scale separable convolutions and Mamba blocks. The multi-scale separable convolution is designed to extract multi-scale fault features from individual sensor signals; by combining depthwise separable convolution with information bottleneck theory, it greatly enhances the capability to capture multi-scale features with only a slight increase in model parameters. However, the global representational power of these multi-scale

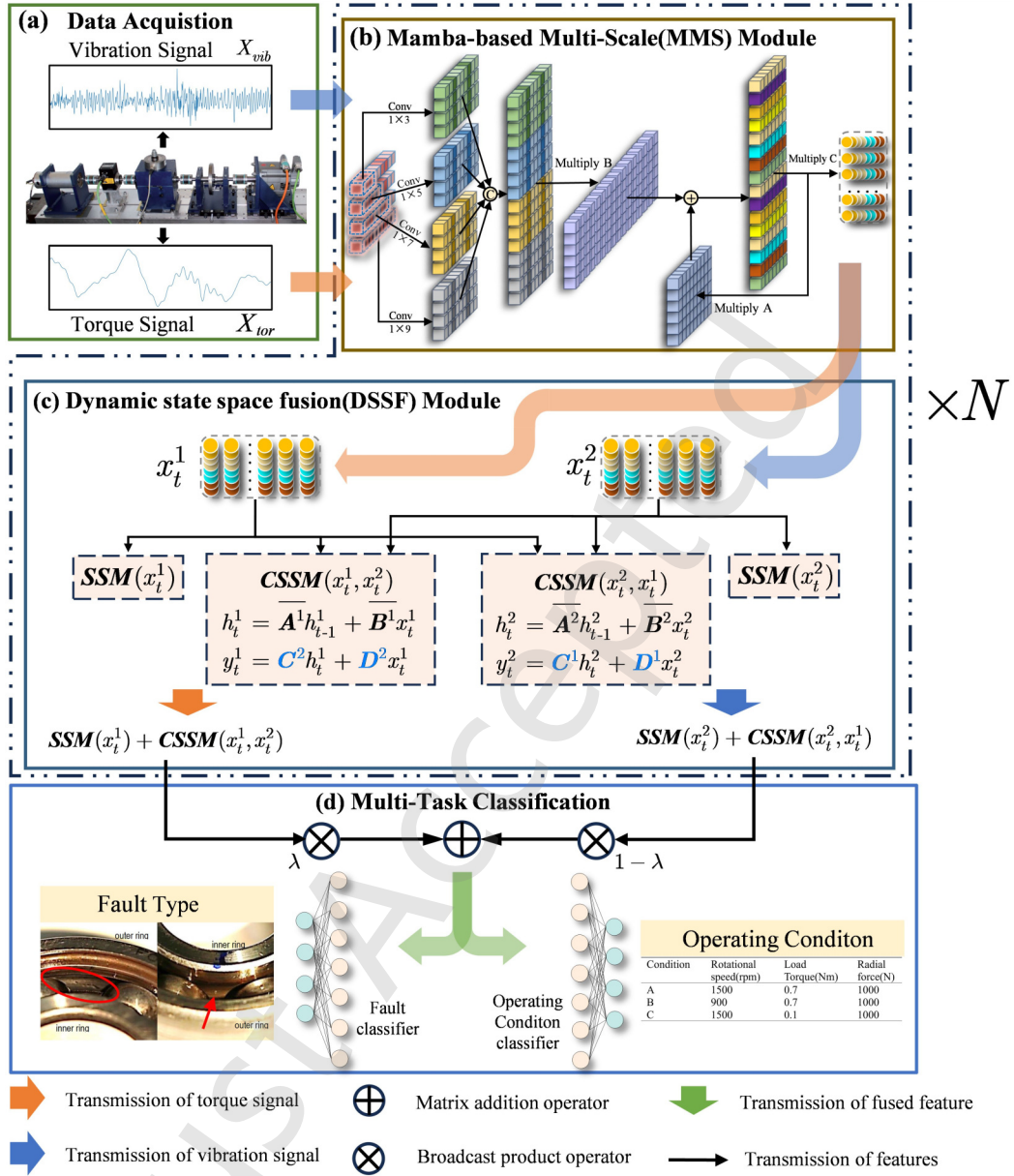


Fig. 1 Structure of the proposed method.

features is limited by the size of the convolutional kernels. Since signals such as bearing vibration and torque span a large temporal range, even large kernels cannot overcome the constraints of a local receptive field. Therefore, a global information modeling mechanism must be integrated into the MMS module. After obtaining multi-scale features, the MMS module employs Mamba's selective hidden-state modeling mechanism to fully capture the global characteristics of the signal.

To address the lack of feature consistency across sensors in existing explicit fusion methods, we introduce the CSSM algorithm within the DSSF

module. The DSSF module takes the multi-scale features output by the MMS module and—via hidden-state feature interaction—produces v^i , with the mapping relationship as follows:

$$v^1, v^2 = \text{DSSF}(u^1, u^2; \Theta_{\text{DSSF}}) \quad (2)$$

where Θ_{DSSF} denotes the parameters to be learned by the DSSF module. By representing cross-sensor information through hidden-state space interaction, it reduces discrepancies among multi-sensor features while fully exploiting their complementary information. It should be noted that a DSSF module is appended after each MMS module, and multiple groups of

MMS-DSSF modules are stacked to build a “local-global feature interaction loop” architecture. Finally, a multi-task classifier synchronously achieves condition classification and fault diagnosis.

3.2 Mamba-based Multi-Scale Module

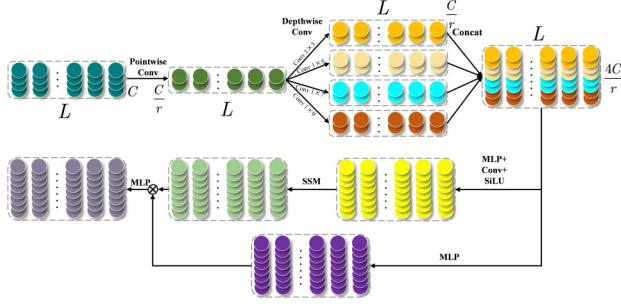


Fig. 2 Mamba-based Multi-Scale module.

The MMS module organically combines multi-scale local feature extraction with global hidden-state modeling to achieve local-global feature extraction in mechanical fault diagnosis. Because local bearing faults produce periodic impulses as rolling elements pass, local feature extraction can detect abnormal amplitudes with high precision, while global feature extraction can model the data distribution of those periodic impulses to ensure that local anomalies are not sporadic. The module deeply integrates a SSM with its selective mechanism under the Mamba framework and incorporates multi-scale depthwise separable convolutions to capture local information. Its overall design inherits the SSM’s advantage in modeling long-range dependencies, effectively compensating for the limited receptive field of conventional convolutions and thus enabling complementary modeling of local details and global correlations under constrained parameters.

The SSM captures long-range dependencies in long sequences via the hidden-state transition matrix and linear recurrence structure, while the Mamba module further introduces a data selection mechanism and adaptive-step discretization to significantly enhance global information modeling efficiency. Meanwhile, multi-scale depthwise separable convolutions perform information compression via pointwise convolution and then capture multi-scale local features through depthwise convolution branches with different receptive fields. Inspired by the Inception architecture and information bottleneck theory, this module emphasizes “preserving salient fault features while suppressing redundant noise.” Through the synergy of these two components, we construct a unified module capable

of jointly extracting multi-scale local details and long-range global dependencies, markedly improving the robustness and accuracy of multi-sensor intelligent mechanical fault diagnosis.

The module is designed as follows: given the input feature of one sensor $X \in \mathbb{R}^{B \times C \times L}$, where B is the batch size, C is the number of channels, and L is the feature length, first apply a pointwise convolution $f^p : \mathbb{R}^{B \times C \times L} \rightarrow \mathbb{R}^{B \times \frac{C}{r} \times L}$ with compression ratio r to achieve feature compression along the channel dimension; its mathematical expression is:

$$X_p = f^p(X) \quad (3)$$

This step compresses redundant information via pointwise convolution—equivalent to a linear projection—in accordance with the core principle of the information bottleneck theory: “preserving salient features and suppressing redundant information.” This facilitates the model’s adaptive suppression of noise while retaining prominent fault characteristics. Subsequently, the module employs four parallel depthwise convolutions with different receptive fields to perform multi-scale feature extraction on X_p , and their output feature maps satisfy:

$$F_k = f_{1 \times k}^d(X_p), k \in \{3, 5, 7, 9\} \quad (4)$$

where $f_{1 \times k}^d$ denotes a depthwise convolution with kernel size $1 \times k$, which generally preserves the number of channels. By using differentiated kernel designs, each branch captures short-range local information (small kernels) and long-range trend features (large kernels), significantly extending the coverage of feature representation. In particular, the parameter count of a depthwise convolution is only $1/k$ of that of a standard convolution, achieving efficient multi-scale modeling. Finally, the outputs of the four branches are concatenated along the channel dimension to produce the multi-scale local feature representation F_{Local} :

$$F_{Local} = \delta(BN(\text{Concat}[F_3, F_5, F_7, F_9])) \quad (5)$$

where $F_{Local} \in \mathbb{R}^{B \times \frac{4C}{r} \times L}$, $\text{Concat}[\cdot]$ denotes concatenation along the channel dimension. δ is the ReLU activation layer, and BN denotes the BatchNorm operation. Here, each sensor’s signal undergoes the above multi-scale feature extraction process, with the local feature representations of sensors 1 and 2 denoted as x_t^1 and x_t^2 , respectively.

Bearing fault signals still exhibit statistical correlations between samples separated by long intervals, such as the periodic impulses of a faulty bearing. To address the insufficient modeling of

long-range dependencies, we construct a dual-stream Mamba architecture that preserves global information via the selective mechanism in the hidden-state space. Its hidden state can be represented by a first-order differential equation describing the evolution of the system's internal state: given the continuous time variable t , the latent state $h^i(t)$ corresponding to the features of the i -th sensor is represented by the state equation:

$$dh^i(t) = (A^i h^i(t) + B^i x^i(t)) dt, i \in \{1, 2\} \quad (6)$$

where the parameters include the hidden-state transition matrix $A^i \in \mathbb{R}^{L \times N}$ and the input projection matrix $B^i \in \mathbb{R}^{B \times C \times N}$. The structured state space model (S4), derived from the vanilla SSM, uses HiPPO for initialization to impose structure on the state matrix A , thereby improving long-range dependency modeling, where $A_{n,k}$ denotes the element in the n -th row and k -th column of A :

$$A_{n,k} = \begin{cases} \sqrt{(2n+1)(2k+1)} & , k < n \\ n+1 & , k = n \\ 0 & , k > n \end{cases} \quad (7)$$

In order to resolve the continuous-time nature of the state equation for compatibility with deep learning algorithms, Mamba discretizes the continuous-form parameters with step size $\Delta = t_k - t_{k-1}$, transitioning to the discrete form:

$$h_t^i = \bar{A}^i h_{t-1}^i + \bar{B}^i x_t^i, \quad (8)$$

$$y_{t,SSM}^i = C^i h_t^i + D^i x_t^i, \quad (9)$$

$$\bar{A}^i = \exp(\Delta A^i), \quad (10)$$

$$\bar{B}^i = (\Delta A^i)^{-1}(\exp(\Delta A^i) - I) \cdot \Delta B^i, \quad (11)$$

$$i \in \{1, 2\} \quad (12)$$

where $\bar{A}^i$ is the discrete hidden-state transition matrix, $\bar{B}^i$ is the discrete input projection matrix, C^i is the output projection matrix, and D^i denotes the residual connection. The dynamic step size Δ induced by the selective mechanism is generated via a differentiable gating mechanism, enabling input-adaptive temporal resolution adjustment. After discretization, efficient computation can be carried out via a linear recurrence method. The detailed operations of the Mamba block are provided in Alg. 1.

Overall, this module employs convolutional kernels of varying sizes to extract an explicit multi-scale set of spatial features with a small parameter budget, and the channel-independence constraint of MMS forces each channel to learn differentiated features,

allowing the model to adaptively match the different periodic patterns of fault signals. Then, through the SSM hidden-state selective mechanism, it effectively captures the long-range spatiotemporal correlations of sensor signals, overcoming the locality limitations of conventional convolutional networks.

Algorithm 1 The process of Mamba Block

Input: $X : (B, C, L)$

Output: $Y : (B, C, L)$

// Generate SSM parameters

1: $x : (B, C, H) \leftarrow \text{Linear}_H(X)$

2: $x' : (B, C, H) \leftarrow \text{SiLU}(\text{Conv1D}(x))$

3: $A : (L, N) \leftarrow \text{Parameter}$

4: $B : (B, C, N) \leftarrow \text{Linear}_N(x')$

5: $C : (B, C, N) \leftarrow \text{Linear}_N(x')$

6: $D : (B, C, H) \leftarrow \text{SiLU}(\text{Linear}_H(X))$

7: $\Delta : (B, C, L) \leftarrow \text{Softplus}(\text{Parameter} +$

$\text{Broadcast}(\text{Linear}_L(x')))$

8: $\bar{A}, \bar{B} : (B, C, L, N) \leftarrow \text{discretize}(\Delta, A, B)$

// Time-varying selectiveSSM

9: $y : (B, C, H) \leftarrow \text{SelectiveSSM}(\bar{A}, \bar{B}, C)(x')$

// Skip connection

10: $y' : (B, C, H) \leftarrow y \otimes D$

11: $Y : (B, C, L) \leftarrow \text{Linear}_L(y')$

12: **return** Y

3.3 Dynamic State Space Fusion Module

In complex scenarios such as multiple operating conditions and multi-level fault severity, single-sensor signals may contain limited information and be heavily affected by noise, leading to suboptimal diagnostic accuracy. Therefore, it is necessary to employ multi-sensor data for collaborative representation of fault features. To achieve multi-sensor feature fusion, one must consider the complex correlations and redundancy inherent in multi-source information from different sensors. Existing methods focus primarily on explicit spatial feature fusion without adequately accounting for the potential heterogeneity among multi-sensor data, which hinders effective alignment during fusion and significantly reduces the consistency of the joint multi-sensor representation. In this paper, we propose the CSSM algorithm, which—combined with two parallel Mamba blocks—performs parameterized cross-sensor feature interaction in the hidden-state space, ensuring thorough alignment and full exploitation of global associations across multi-source information.

As shown in Fig. 3, we introduce the CSSM algorithm for exchanging hidden-state information on top of two parallel Mamba blocks. Given a pair of

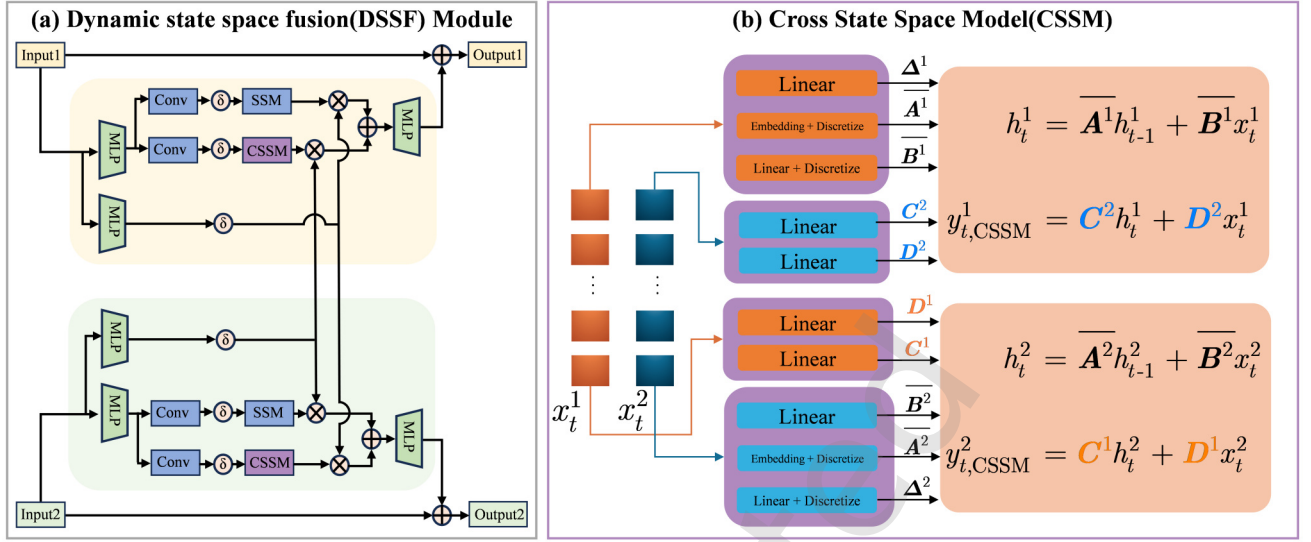


Fig. 3 Dynamic state space fusion module.

multi-sensor local features at the same layer, x_t^1, x_t^2 , the corresponding SSM parameters $(\overline{A}^i, \overline{B}^i, C^i, D^i)$ are generated for each input following the steps of Alg. 1. Similarly, each hidden state is first updated as:

$$h_t^i = \overline{A}^i h_{t-1}^i + \overline{B}^i x_t^i, i \in \{1, 2\} \quad (13)$$

Up to this point, the forward propagation processes of SSM and CSSM remain identical; their differences can be described by the following equation:

$$y_{t,SSM}^1 = C^1 h_t^1 + D^1 x_t^1 \quad (14)$$

$$y_{t,SSM}^2 = C^2 h_t^2 + D^2 x_t^2, \quad (15)$$

$$y_{t,CSSM}^1 = C^1 h_t^1 + D^2 x_t^1, \quad (16)$$

$$y_{t,CSSM}^2 = C^2 h_t^2 + D^1 x_t^2 \quad (17)$$

where $y_{t,SSM}^1, y_{t,SSM}^2$ correspond to the outputs of the standard Mamba blocks for x_t^1, x_t^2 , respectively. CSSM, by performing parameter cross in the hidden-state space, realizes feature interaction of the inputs x_t^1, x_t^2 within that space, thereby implicitly learning the collaborative mapping relationships of cross-sensor features. It ultimately produces outputs $y_{t,CSSM}^1, y_{t,CSSM}^2$. The visualized computation process of CSSM is shown on the right side of Fig. 3.

The outputs of this module are then enhanced via element-wise addition and a fully connected layer as follows:

$$F_{Global}^1 = \text{Linear}(y_{t,SSM}^1 + y_{t,CSSM}^1) + x_t^1, \quad (18)$$

$$F_{Global}^2 = \text{Linear}(y_{t,SSM}^2 + y_{t,CSSM}^2) + x_t^2 \quad (19)$$

where $F_{Global}^1, F_{Global}^2$ denote the global features obtained from the inputs of sensors 1 and 2 via hidden-state space feature interaction.

By stacking MMS and DSSF modules, the “local-global feature interaction loop” is realized. Assuming there are N stacked MMS and DSSF modules, this loop can be described as follows:

$$F_{Global,i}^1, F_{Global,i}^2 = \text{DSSF}(\text{MMS}(F_{Global,i-1}^1, F_{Global,i-1}^2)), \quad (20)$$

$$i \in \{1, 2, \dots, N\}$$

where $F_{Global,i}^1, F_{Global,i}^2$ are the outputs of the i -th MMS and DSSF modules; when $i = 1$, their inputs are initialized to the original sensor signals. After N iterations, a learnable parameter $\lambda \in (0, 1)$ is used to adaptively weight and fuse the multi-sensor signals:

$$F_{Fused} = \lambda F_{Global,N}^1 + (1 - \lambda) F_{Global,N}^2 \quad (21)$$

where F_{Fused} denotes the fused feature, which is then fed into two classifier networks of identical architecture but with independent parameters to perform multi-task classification. The classifier networks can be described as follows:

$$\text{Output} = \text{Linear}(GAP(f_{1 \times 3}(F_{Fused}))) \quad (22)$$

where $f_{1 \times 3}$ is a convolutional layer with kernel size 3, and GAP denotes global average pooling. This study examines the network’s ability to perform multiple tasks simultaneously, with each task framed as a multi-class classification problem—for example, fault-type classification or operating-condition classification. For t tasks, the multi-task classification loss function is

defined as:

$$Loss(X_{1:s}, Y_{1:t}) = \sum_{i=1}^t w_i loss_i(X_{1:s}, Y_i) \quad (23)$$

$$loss_i = - \sum_{j=1}^n y_j \log(\hat{y}_j) \quad (24)$$

where $X_{1:s}$ denotes the signal samples from s sensors, and $Y_{1:t}$ denotes the task-specific labels. Considering that directly summing the loss functions of multiple tasks may cause imbalance, we introduce the hyperparameter w_i to balance each task's weight. Each task's loss function is defined as cross-entropy, where n is the total number of labels for the i -th multi-class task, $\hat{y}_j$ is the predicted class probability, and $loss_i$ is the loss value for the i -th multi-class task.

In summary, the DSSF module proposed in this section, through the parameterized interaction mechanism of the dual-stream Mamba blocks, effectively avoids the dimensionality explosion caused by traditional explicit feature concatenation while enhancing the consistency of multi-sensor features. Moreover, the stacking mechanism of the local-global feature interaction loop progressively eliminates inter-sensor information redundancy via multi-layer feature extraction, achieving gradual joint optimization of multi-source information representation; finally, by innovatively introducing a learnable adaptive weighting fusion coefficient λ , the model is endowed with adaptive dynamic sensing capability for sensor information.

4 Experimental Results and Analysis

4.1 Experimental Setup

In order to validate the effectiveness of the proposed model, this section presents an experimental study. The average prediction accuracy is selected as the final metric for all comparisons, computed as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (25)$$

where TP , FP , FN , and TN denote the numbers of true positives, false positives, false negatives, and true negatives, respectively. To better simulate the noise interference encountered by rolling bearings in actual operation, Gaussian noise is added to the original signal. The signal-to-noise ratio (SNR) is defined as follows:

$$SNR_{dB} = 10 \log_{10}(P_{\text{signal}}/P_{\text{noise}}) \quad (26)$$

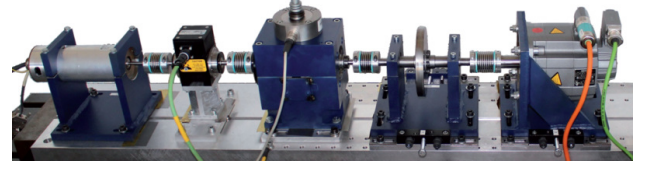


Fig. 4 Paderborn dataset test rig.

Table 1 The health status of the bearings in the PU dataset used in the experiment.

Damages type	Damages level	Labels
Healthy	\	0
Outer ring fault	1	1
Outer ring fault	2	2
Inner ring fault	1	3
Inner ring fault	2	4

Table 2 The operating conditions of the bearings in the PU dataset used in the experiment.

Condition Labels	Rotational speed(rpm)	Load Torque(N·m)	Radial force(N)
A	1500	0.7	1000
B	900	0.7	1000
C	1500	0.1	1000

where P_{signal} and P_{noise} denote the power of the signal and the noise, respectively.

The model was trained and evaluated on a system with an Intel i9-10940X CPU and an RTX 3090 GPU, running Ubuntu 20.04, Python 3.9.19, and PyTorch 1.31.1. The proposed network stacks five layers of the “local-global feature interaction loop.” The base learning rate was set to 0.001, the batch size to 64, and the optimizer was Adam with weight decay = $1e - 5$ and momentum parameters $\beta_1, \beta_2 = 0.9, 0.999$. Five-fold cross-validation was adopted, and training was conducted for a total of 100 epochs.

4.2 Dataset Description

4.2.1 PU Bearing Fault Diagnosis Datasets

One of the experimental datasets used in this study is the publicly available rolling bearing condition monitoring dataset released by the Chair of Design and Drive Technology at Paderborn University, Germany. This dataset was generated through systematic experimental design, encompassing various bearing damage types and operating conditions, and has been widely used in data-driven fault diagnosis research. As depicted in Fig. 4, the test bench test

rig comprises several modules: an electric motor, a torque-measurement shaft, a rolling bearing test module, a flywheel, and a load motor.

The test bearings (model 6203) are integrated into the bearing test module to simulate bearing faults under actual operating conditions through artificial damage generation and accelerated lifetime tests. Artificial damage, including single-point defects on the inner and outer rings, is introduced via processes such as electrical discharge machining (EDM), drilling, and electric engraving. Real damage, such as fatigue pitting and plastic deformation, is induced by accelerated lifetime tests under high-load and low-lubrication conditions. The dataset comprises 32 bearing samples, including healthy states (6 groups), artificial damage (12 groups), and real damage (14 groups), with fault types covering inner ring faults, outer ring faults, and others. The bearing box's acceleration is gauged using a piezoelectric accelerometer (model 336C04, PCB Piezotronics, Inc.), paired with a charge amplifier (model 5015A, Kistler Group) and a 30 kHz low-pass filter. This assembly is located atop the rolling bearing module. The acquired signals are digitized and archived synchronously with the MCS at a 64 kHz sampling rate. Meanwhile, torque signals are obtained using a torque transducer (Torque Transducer model 305, Magtrol, 2 Nm) with a sampling frequency of 4 kHz.

In the experiments conducted, five bearing health states and three operating conditions were selected for validation. To ensure experimental accuracy, all faulty bearings used were those with actual faults: including healthy bearings, two severity levels of outer-ring faults, and two severity levels of inner-ring faults. These bearings were tested under three different operating conditions, labeled A, B, and C, with bearing state labels from 0 to 4. The health states of the bearings used in the experiments are listed in Table 1. Damage levels are determined by the percentage of defect length relative to the pitch diameter, where level-1 damage ranges from 0-2%, and level-2 damage ranges from 2-5%. The operating conditions used in the experiments are listed in Table 2.

A total of 9000 samples were prepared for this study, each comprising a sliced vibration signal and a torque signal. The vibration signal consists of 2048 data points; since the torque signal's sampling rate is 1/16 of that of the vibration signal, each sample's torque signal contains only 128 data points. To align multi-sensor data in the time dimension, the lower-rate data

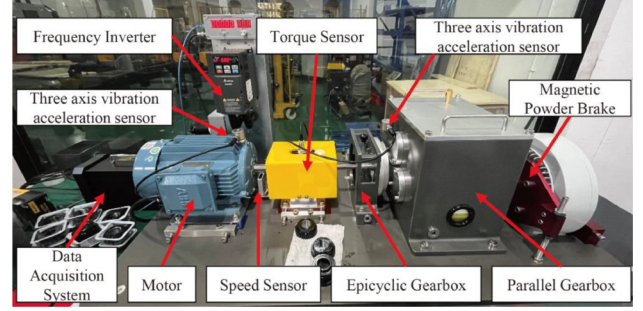


Fig. 5 MCC5-THU dataset test rig.

Table 3 The health status of the gearbox in the MCC5-THU dataset used in the experiment.

Damages type	Labels
Healthy	0
Miss teeth	1
Teeth crack	2
Gear wear	3
Teeth break	4
Gear pitting	5

Table 4 The operating conditions of the gearbox in the MCC5-THU dataset used in the experiment.

Condition Labels	Rotational speed(rpm)
A	500-1000
B	1500-2000
C	2500-3000

are upsampled so that each sensor has the same number of data points. Assuming an input signal of length N , denoted x , is upsampled by a factor of k , the process can be described as follows:

$$X = FFT(x) \quad (27)$$

$$Y = \text{zero_pad}(X) \quad (28)$$

$$y = k \cdot FFT^{-1}(Y) \quad (29)$$

where $FFT(\cdot)$ and $FFT^{-1}(\cdot)$ denote the Fourier transform and inverse Fourier transform, respectively, and $\text{zero_pad}(\cdot)$ denotes zero-padding in the upsampled portion. This paper also employs a simple yet effective method for data normalization, computed as shown in the following equation:

$$\hat{x} = \frac{x - \mu}{\sigma} \quad (30)$$

where x denotes the input, μ represents the mean of the data, and σ represents the standard deviation of the data.

4.2.2 MCC5-THU Gearbox Fault Diagnosis Datasets

To achieve a more comprehensive validation, particularly regarding the generalization capability to

other types of rotating machinery, an additional and more complex open-source gearbox dataset (MCC5-THU gearbox dataset) was employed to further evaluate the advantages of the proposed MHSNet. This dataset, provided by MCC5 Group Shanghai Co., Ltd. and Tsinghua University, contains vibration signals and torque signals recorded from real gearbox faults under different rotational speeds. The experimental setup for the dataset is shown in Fig. 5, which consists of a 2.2 kW three-phase asynchronous motor, a torque sensor, a two-stage parallel gearbox, a magnetic powder brake acting as a torque generator, and a measurement and control system. The actual torque endured by the gearbox input shaft can be measured by a torque sensor (model S2001). The test rig is equipped with two three-axis vibration acceleration sensors (model TES001V) were used to measure motor and gearbox intermediate shaft triaxial vibrations along the x-, y-, and z-axes at a sampling frequency of 12.8 kHz.

Several experimental configurations with different fault types and operating conditions were implemented on the gearbox. In this study, all gear fault types and variable-speed conditions were selected for fault diagnosis experiments, comprising six gear health states and three operating conditions. These include the healthy state, missing teeth, tooth crack, gear wear, tooth breakage, and gear pitting, with each tested under three different variable-speed conditions. The health states and operating conditions of the bearings used in the experiments are summarized in Table 3 and Table 4, respectively. A detailed description of the dataset can be found in [57].

In total, 4500 samples were prepared for the experiments. Each sample consists of segmented single-channel vibration and single-channel torque signals, with both signals containing 2048 data points. Prior to model training, all signals were normalized using the same procedure as that applied in the PU dataset experiments, ensuring comparability between datasets.

4.3 Comparison of Experimental Results

4.3.1 Comparative Experiments on The PU Dataset

In this section, we compare the performance of MHSNet with six other fault diagnosis networks under no noise and three different noise levels (-1 dB, -3 dB, and -5 dB). Among them, HCNN-1D [47], MTA [30] and MT-ConvFormer [44] are single-modality multi-

task models; M2FN [29], MICN [58] and MRCFN [59] are single-task models based on heterogeneous multi-sensor signals, where the heterogeneous multi-sensor signal inputs include vibration and torque signals, while the single-modality input consists only of vibration signals. The fault types diagnosis task (FTD-task) and the operating condition classification task (OCC-task) were performed.

Table 5 presents the average diagnostic accuracy, standard deviation, and the variation of diagnostic accuracy with noise for each network using five-fold cross-validation. Fig. 6 shows a more intuitive display of the accuracy of each network under different noise environments. Under no noise conditions, all seven models perform well in fault diagnosis tasks. When noise with a signal-to-noise ratio (SNR) of -1 dB is introduced, the fault diagnosis accuracy of MHSNet barely decreases, and the operating condition classification accuracy slightly decreases but remains stable at 94.44%. When the noise increases to -3 dB, MHSNet's fault diagnosis accuracy remains at 99.39%, and the operating condition classification accuracy reaches 92.56%.

When the noise level increases to -5 dB, the fault diagnosis performance of the three single-task models, HCNN-1D, MTA and MT-ConvFormer, deteriorates, while the single-task models based on heterogeneous multi-sensor signals, M2FN, MRCFN and MICN, along with our model, achieve relatively higher fault diagnosis and operating condition classification accuracy compared to traditional single-modality methods. This demonstrates that using multi-sensor data for collaborative feature representation avoids information loss or noise interference issues that might arise in complex conditions with single-sensor signals. However, the fault diagnosis and operating condition classification accuracy of M2FN and MRCFN also show a significant decrease. In contrast, MHSNet achieves a fault diagnosis accuracy of 98.24%, almost unaffected by noise, and despite a slight drop to 90.89% in the operating condition classification task, MHSNet outperforms M2FN, MRCFN and MICN in operating condition classification accuracy (77.57%, 85.32% and 90.33%, respectively), improving by 13.32%, 5.57% and 0.56%, respectively. This indicates that through the two parallel Mamba blocks, which implement parameterized interaction of cross-sensor features in the hidden-state space, MHSNet ensures effective alignment of multi-source information and

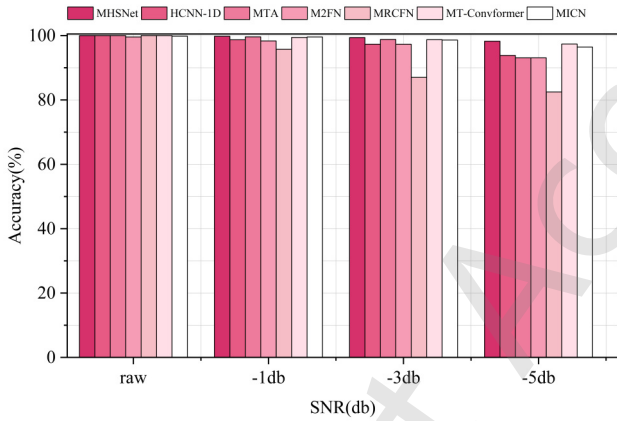
Table 5 Accuracy of different models in the PU dataset under different SNR levels.

(a) Compared with single-modal multi-task models

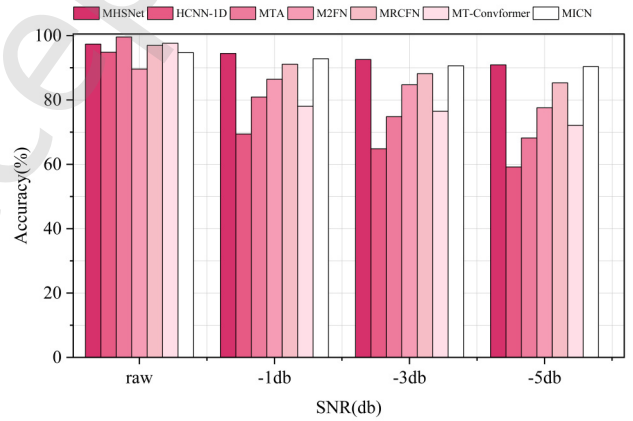
SNR(dB)	MHSNet		HCNN-1D		MTA		MT-Convformer	
	FTD-task	OCC-task	FTD-task	OCC-task	FTD-task	OCC-task	FTD-task	OCC-task
raw	100.00±0.00	97.34±1.77	100.00±0.00	94.83±1.49	100.00±0.00	99.56±0.12	100.00±0.00	97.64±1.93
-1dB	99.79±0.14	94.44±0.54	98.70±0.32	69.43±1.67	99.61±0.30	80.89±0.51	99.43±0.11	78.03±0.70
-3dB	99.39±0.62	92.56±1.11	97.32±0.84	64.83±2.29	98.81±0.38	74.81±1.98	98.74±0.17	76.49±0.70
-5dB	98.24±2.11	90.89±1.17	93.82±0.49	59.18±1.26	93.09±7.60	68.17±2.90	97.39±0.56	72.14±1.20

(b) Compared with single-task models based on heterogeneous multi-sensor signals

SNR(dB)	MHSNet		M2FN		MRCFN		MICN	
	FTD-task	OCC-task	FTD-task	OCC-task	FTD-task	OCC-task	FTD-task	OCC-task
raw	100.00±0.00	97.34±1.77	99.56±0.17	89.61±1.57	100.00±0.00	96.96±0.20	99.82±0.25	94.71±4.57
-1dB	99.79±0.14	94.44±0.54	98.33±0.48	86.43±0.84	95.74±2.25	91.11±1.14	99.58±0.57	92.81±3.10
-3dB	99.39±0.62	92.56±1.11	97.31±0.51	84.74±2.22	87.00±3.30	88.14±1.74	98.61±1.23	90.62±6.60
-5dB	98.24±2.11	90.89±1.17	93.10±3.71	77.57±12.95	82.46±4.33	85.32±2.35	96.43±5.86	90.33±1.43



(a)



(b)

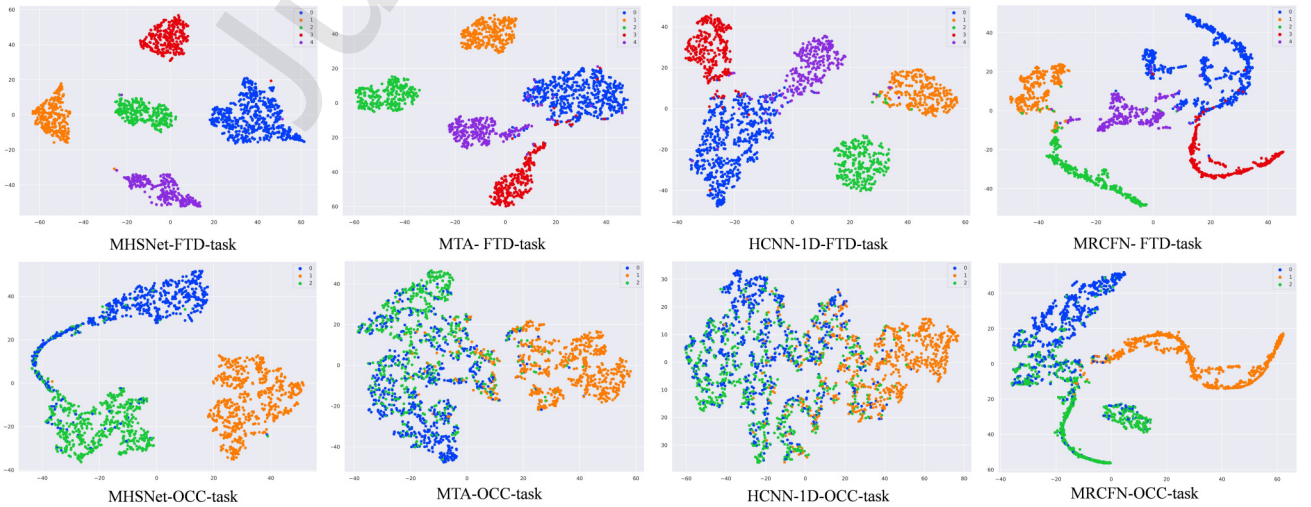
Fig. 6 Accuracy of different models in the PU dataset under different SNRs. (a) FTD-task. (b) OCC-task.**Fig. 7** Feature visualization via t-SNE under -5 -dB noise interference.

Table 6 Accuracy results of different models in various unbalanced operating conditions on the PU dataset under -5 dB.

Data Volume(%)	MHSNet		M2FN		MRCFN	
	FTD-task	OCC-task	FTD-task	OCC-task	FTD-task	OCC-task
100,100,100	98.24$\pm$2.11	90.89$\pm$1.17	93.10 $\pm$ 3.71	77.57 $\pm$ 12.95	82.46 $\pm$ 4.33	85.32 $\pm$ 2.35
80,10,10	92.63$\pm$5.41	87.23$\pm$3.00	91.20 $\pm$ 1.98	80.77 $\pm$ 1.07	73.70 $\pm$ 9.23	81.63 $\pm$ 3.93
60,20,20	93.80$\pm$6.49	80.00$\pm$1.33	90.40 $\pm$ 1.25	70.33 $\pm$ 3.55	76.70 $\pm$ 6.33	67.13 $\pm$ 6.04
40,60,60	95.47$\pm$2.97	70.37$\pm$10.74	91.10 $\pm$ 1.64	66.00 $\pm$ 3.07	70.63 $\pm$ 5.57	60.43 $\pm$ 9.50

further improves the model's performance in multi-sensor signal data fusion.

Furthermore, to evaluate the quality of the extracted classification features, we used t-distributed Stochastic Neighbor Embedding (t-SNE) for dimensionality reduction and feature visualization, as shown in Fig. 7. From the t-SNE results, it can be observed that our model better extracts features. These experimental comparisons validate the advantages of MHSNet in bearing fault diagnosis. MHSNet uses the Mamba module for multi-scale feature extraction and performs multi-sensor feature fusion through the DSSF module, enabling it to better capture feature information and temporal relationships from vibration signals, thereby effectively enhancing fault diagnosis robustness and accuracy.

Finally, Table 6 shows the accuracy of three models (MHSNet, M2FN, MRCFN) under different data imbalance conditions at -5 dB SNR. The dataset is divided into four different imbalance situations: a balanced dataset (100,100,100), where the data proportions for each category are equal, and three imbalanced datasets (80,10,10; 60,20,20; 40,60,60), where the percentages of the three operating condition samples are as shown in Table 2. From Table 6, it can be seen that MHSNet performs excellently under data imbalance conditions, particularly in the FTD-task and OCC-task, consistently maintaining high accuracy and being more robust than the other two models (M2FN and MRCFN), effectively handling data imbalance under high noise conditions. The comparison shows that MHSNet has clear advantages in multi-sensor data fusion and inter-task information sharing, especially when dealing with data imbalance. Its performance remains stable and superior to other single-modality and heterogeneous multi-sensor signal-based models.

4.3.2 Comparative Experiments on The MCC5-THU Dataset

To further validate the performance of MHSNet, this section conducts comparative experiments on the

MCC5-THU dataset against six other fault diagnosis networks. Similarly, tests are performed under noise-free conditions and three different noise levels (-1 dB, -3 dB, and -5 dB). For heterogeneous multi-sensor input, both vibration and torque signals are included, while the single-modality input contains only vibration signals. The experiments are conducted for the Fault Type Diagnosis task (FTD-task) and the Operating Condition Classification task (OCC-task).

Table 7 presents the average diagnostic accuracy, standard deviation, and variation in accuracy with noise for each network using five-fold cross-validation. Under noise-free conditions, MHSNet achieves the highest FTD-task accuracy of 95.96% and a perfect 100% accuracy in the OCC-task. When subjected to noise with a signal-to-noise ratio (SNR) of -1 dB, MHSNet still maintains a stable FTD-task accuracy of 79.64% $\pm$ 2.15, outperforming other models, while its OCC-task accuracy remains at a high level of 99.84%.

When the noise level increases to -3 dB, the performance of almost all models begins to degrade to varying degrees. Nevertheless, MHSNet continues to lead, with an FTD-task accuracy of 78.04% and an OCC-task accuracy of 99.36%, both the highest among all models, demonstrating exceptional robustness. MHSNet's strong performance in this moderate-noise environment is mainly attributed to its multi-task collaborative learning mechanism. By jointly performing fault diagnosis and operating condition classification, the model can extract more discriminative and complementary features, effectively enhancing its anti-interference capability and reducing the performance drop far more than other models.

Under the most challenging environment with a noise level of -5 dB, the fault diagnosis performance of single-modality models such as HCNN-1D, MTA, and MT-Convformer declines significantly. In contrast, MHSNet still achieves the highest FTD-task accuracy of 74.40% with minimal standard deviation, indicating the stability of its diagnostic results. In the OCC-task,

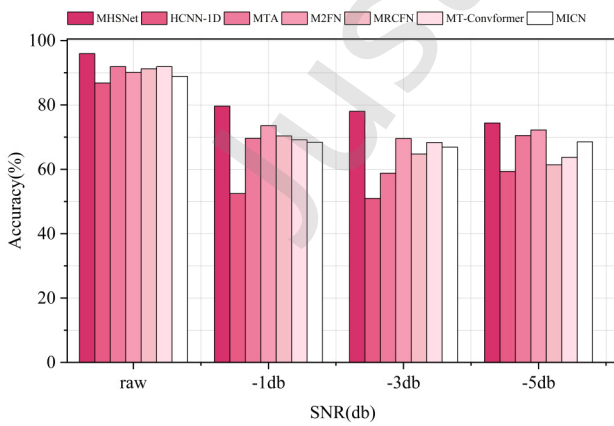
Table 7 Accuracy of different models in the MCC5-THU dataset under different SNR levels.

(a) Compared with single-modal multi-task models

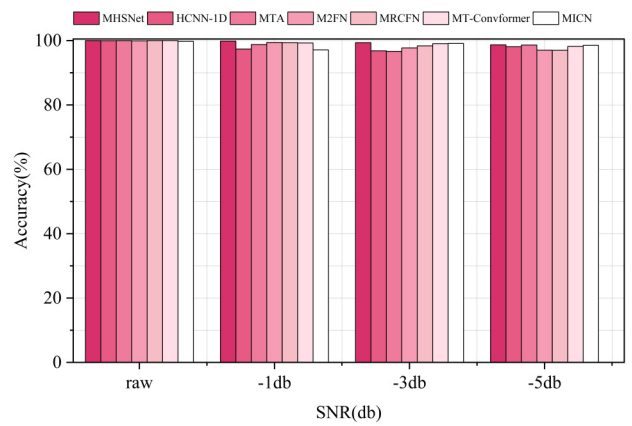
SNR(dB)	MHSNet		HCNN-1D		MTA		MT-Convformer	
	FTD-task	OCC-task	FTD-task	OCC-task	FTD-task	OCC-task	FTD-task	OCC-task
raw	95.96±2.52	100.00±0.00	86.84±1.11	99.98±0.04	91.91±0.98	100.00±0.00	91.91±1.33	100.00±0.00
-1dB	79.64±2.15	99.84±0.05	52.47±1.46	97.38±0.41	69.60±1.83	98.76±0.50	69.18±1.48	99.27±0.27
-3dB	78.04±2.99	99.36±0.91	50.96±2.01	96.82±0.79	58.80±2.42	96.67±1.20	68.33±1.46	99.04±0.48
-5dB	74.40±0.92	98.67±0.62	59.36±1.89	98.09±0.77	70.51±1.30	98.62±0.60	63.71±2.28	98.18±0.54

(b) Compared with single-task models based on heterogeneous multi-sensor signals

SNR(dB)	MHSNet		M2FN		MRCFN		MICN	
	FTD-task	OCC-task	FTD-task	OCC-task	FTD-task	OCC-task	FTD-task	OCC-task
raw	95.96±2.52	100.00±0.00	90.13±1.26	99.91±0.08	91.24±4.06	100.00±0.00	88.89±1.66	99.76±0.33
-1dB	79.64±2.15	99.84±0.05	73.56±3.49	99.42±0.23	70.38±1.01	99.36±0.15	68.40±2.35	97.09±3.27
-3dB	78.04±2.99	99.36±0.91	69.56±2.27	97.73±0.74	64.76±5.43	98.38±0.73	66.89±2.22	99.11±0.69
-5dB	74.40±0.92	98.67±0.62	72.22±2.59	97.00±0.88	61.38±4.95	96.98±0.79	68.51±2.16	98.51±0.64



(a)



(b)

Fig. 8 Accuracy of different models in the MCC5-THU dataset under different SNRs. (a) FTD-task. (b) OCC-task.

Table 8 Accuracy results in task ablation experiments for each task on the PU dataset.

SNR(dB)	MHSNet		MHSNet-FTD		MHSNet-OCC	
	FTD-task	OCC-task	FTD-task	OCC-task	FTD-task	OCC-task
raw	100.00±0.00	97.34±1.77	99.99±0.02	N/A	N/A	95.66±2.59
-1dB	99.79±0.14	94.44±0.54	99.74±0.19	N/A	N/A	87.33±5.24
-3dB	99.39±0.62	92.56±1.11	91.40±16.56	N/A	N/A	86.31±4.44
-5dB	98.24±2.11	90.89±1.17	89.33±11.94	N/A	N/A	81.47±5.66

MHSNet likewise delivers the best performance among all models. Its outstanding performance in such harsh conditions is primarily due to its core selective hidden state modeling mechanism, which plays a crucial role under strong noise by effectively distinguishing periodic impact features from the overwhelming noise in the signals. Combined with multi-scale convolutional units, it ensures simultaneous capture of local details and global dependencies. Moreover, the dynamically designed state-space fusion module, tailored to address heterogeneity, performs parameterized interaction and adaptive alignment of multi-sensor noisy signals, effectively suppressing noise interference and ensuring deep fusion of valid information.

In summary, the comparative experiments on the MCC5-THU dataset provide strong evidence that MHSNet exhibits significant advantages in multi-sensor collaborative representation, global-local feature fusion, and noise suppression. It consistently achieves the best fault diagnosis accuracy and operating condition classification performance, surpassing traditional single-modality and heterogeneous multi-sensor signal diagnosis networks.

4.4 Ablation Studies

First, to explore the impact of multi-task learning on the FTD-task and OCC-task, this experiment compares models that only handle FTD-task and only handle OCC-task. An ablation study was conducted under

the same conditions as the comparative experiments on the PU dataset, and the results are shown in Table 8. From Table 8, it can be seen that as noise increases, compared to the FTD single-task, our model with multi-task learning achieves improvements of 0.01%, 0.05%, 7.99%, and 8.91%; compared to the OCC single-task, our model with multi-task learning achieves improvements of 1.68%, 7.11%, 5.09%, and 9.42% in the OCC task. The results indicate that, especially under noise interference, using MHSNet to jointly train multiple fault diagnosis tasks allows task-specific features to contain useful information that can be leveraged by other tasks, thus improving performance.

Next, in the comparison experiments where the Mamba fusion module is removed and only single-modality feature concatenation (V1) is used, and where multi-scale convolutions are replaced by standard convolutions (V2), the results in Table 9 show that the full model achieves the following performance: In the FTD-task, under no noise interference, -1 dB, -3 dB, and -5 dB noise, it achieves 100.00%, 99.79%, 99.39%, and 98.24%; in the OCC-task, it achieves 97.34%, 94.44%, 92.56%, and 90.89%. However, in the V1 model, the FTD accuracy drops to 99.59%, 97.57%, 93.46%, and 96.99% under the corresponding conditions, and in the OCC-task, it achieves 94.09%, 90.07%, 91.26%, and 89.89% under no noise and -1 dB, -3 dB, -5 dB noise, respectively. In the V2 model, the FTD accuracy is 99.98%, 99.12%, 98.74%, and 98.34%, and in the OCC-task, it achieves 97.26%, 92.17%, 92.83%, and 90.61% under the same noise conditions. Finally, based on the comparison of experimental results, histograms are plotted, as shown in Fig. 10, and t-SNE visualization is shown in Fig. 9. It can be observed that although in some noise conditions or specific metrics, V2 occasionally shows values close to the full model, overall, the full model outperforms the other two variants in most cases. This indicates that the multi-scale convolution effectively captures fault signal details and trend features under different receptive

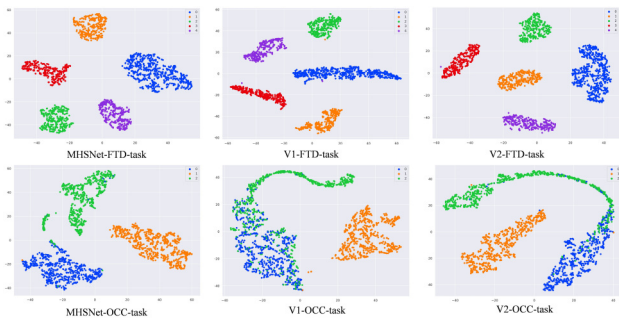
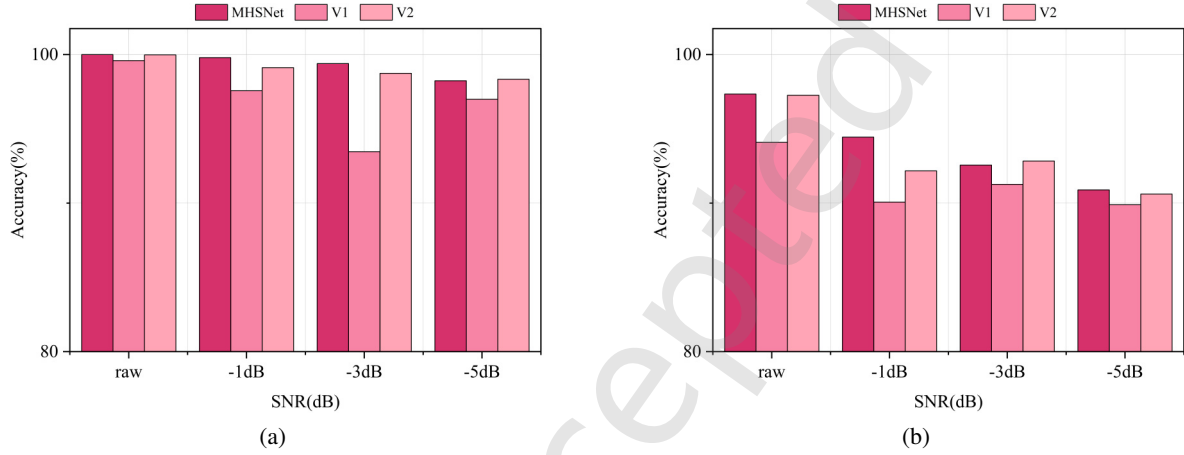


Fig. 9 Feature visualization via t-SNE under -1 -dB noise interference.

Table 9 Accuracy results in module ablation experiment for each module on the PU dataset.

SNR(dB)	MHSNet		V1		V2	
	FTD-task	OCC-task	FTD-task	OCC-task	FTD-task	OCC-task
raw	100.00±0.00	97.34±1.77	99.59±0.79	94.09±5.59	99.98±0.03	97.26±1.60
-1dB	99.79±0.14	94.44±0.54	97.57±4.34	90.07±6.88	99.12±0.50	92.17±2.25
-3dB	99.39±0.62	92.56±1.11	93.46±8.64	91.26±2.62	98.74±1.91	92.83±1.83
-5dB	98.24±2.11	90.89±1.17	96.99±4.47	89.89±1.02	98.34±1.23	90.61±2.11

**Fig. 10** Accuracy results in module ablation experiments for each module under different SNRs on the PU dataset. (a) FTD-task. (b) OCC-task.

fields, providing the model with a more comprehensive representation ability in the presence of local periodic shocks and multi-level fault modes. The Mamba module, through its fusion mechanism in the hidden-state space, enables adaptive interaction and alignment of multi-source features and provides global modeling capabilities, thus making the model more robust and accurate under complex noise interference and multi-condition scenarios. Although standard convolutions might show “unexpected” higher values in certain noise conditions or feature distributions due to their simpler network structure, their generalization ability in most conditions is usually insufficient, especially in scenarios with high noise or multiple fault modes, where their performance still lags behind the overall performance of multi-scale convolutions.

Finally, we compared the experimental results under single-modality, using only vibration modality or only torque modality. As shown in Table 10, it can be seen that as noise increases, if only the vibration modality is used, it performs better in detecting the FTD task, while using only the torque modality performs better in detecting the OCC task, though the FTD task performance is poor. However, when both vibration and torque modality data are used as input, our model

performs well in both the FTD and OCC tasks.

Table 10 Accuracy results of ablation experiments for multimodal and unimodal data on the PU dataset.

SNR(dB)	MHSNet		MHSNet-vibration		MHSNet-torque	
	FTD-task	OCC-task	FTD-task	OCC-task	FTD-task	OCC-task
raw	100.00±0.00	97.34±1.77	99.68±0.56	88.52±4.40	95.48±2.60	96.36±1.88
-1dB	99.79±0.14	94.44±0.54	92.77±11.72	72.00±6.54	93.97±1.03	86.82±5.96
-3dB	99.39±0.62	92.56±1.11	96.72±1.22	66.92±5.52	87.77±5.34	88.10±2.20
-5dB	98.24±2.11	90.89±1.17	90.51±7.04	65.27±1.76	86.38±2.56	84.30±3.83

5 Conclusion

This paper proposes a multi-sensor and multi-task fault diagnosis method based on the Multi-Scale Hidden State Interaction Network (MHSNet). By integrating the advantages of deep separable convolution and hidden SSMs, it achieves the joint modeling of local details and global dependency information. Specifically, MHSNet uses multi-scale convolution units to capture local fault features in signals such as bearing vibration and torque under different receptive fields, and with the selective hidden state modeling mechanism of the Mamba module, it successfully overcomes the limitations of traditional convolution networks in global information extraction, thereby effectively characterizing the periodic shocks and long-range dependencies in the signals. In terms of multi-sensor data fusion, the proposed Dynamic State Space Fusion Module (DSSF) achieves adaptive alignment and feature interaction of different sensor data through parameterized interactions in the hidden state space, significantly alleviating the distribution differences and redundancy issues between multi-source information. Combined with the multi-task learning framework, this method further exploits complementary information between tasks, resulting in a significant improvement in overall fault diagnosis and condition classification accuracy under data imbalance and noise interference conditions. Extensive experimental results demonstrate that MHSNet maintains high diagnostic accuracy and robustness under both no noise and various noise levels, especially outperforming traditional unimodal models and other heterogeneous multi-sensor signal-based models in terms of fault diagnosis and condition classification performance under high noise conditions.

Overall, this research makes significant progress in the field of heterogeneous multi-sensor signal data fusion and fault diagnosis modeling. It not only improves diagnostic accuracy under multiple fault types and operating conditions but also provides an efficient and robust solution for intelligent monitoring

of complex industrial systems. In the future, further exploration of deeper cross-modal feature interaction mechanisms for heterogeneous multi-sensor signals, extending the model's application to other complex mechanical systems, and integrating advanced methods such as self-attention will provide broader research space and practical application potential for equipment health monitoring and fault prediction.

Acknowledgment

This research is supported by the National Natural Science Foundation of China (Grant No. 52302444, 52202424), the National Natural Science Foundation of China (No. U2468207), the Sichuan Science and Technology Program (No. 2024ZHCG0166), the Natural Science Foundation of Sichuan Province (Grant No. 2024NSFSC0908, 2025ZNSFSC0400), the National Key R&D Program of China (Grant No. 2021YFB3400704, 2021YFB3400704-02), Independent Project of State Key Laboratory of Rail Transit Vehicle System (Grant No. 2024RVL-T05), Fundamental Research Funds for the Central Universities (Grant No. 2682023GF012), Fundamental Research Funds for the Central Universities (Grant No. A1420502052501-64), Sichuan Science and Technology Program (Grant No. 2022JDRC0067).

References

- [1] X. Wang, H. Jiang, T. Zeng, and Y. Dong, "An adaptive fused domain-cycling variational generative adversarial network for machine fault diagnosis under data scarcity," *Information Fusion*, p. 103616, 2025.
- [2] C. Yi, L. Ran, J. Tang, Q. Zhou, and L. Zhou, "Harmonic spectral correlated kurtosis and an adaptive matching extraction strategy of multi-fault features for rotating machinery," *Structural Health Monitoring*, vol. 23, no. 2, pp. 1245–1262, 2024.
- [3] B. Chen, Y. Cheng, P. Allen, S. Wang, F. Gu, W. Zhang, and A. D. Ball, "A product envelope spectrum generated from spectral correlation/coherence for railway axle-box bearing

- fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 225, p. 112262, 2025.
- [4] Q. Wang, R. Huang, J. Xiong, J. Yang, X. Dong, Y. Wu, Y. Wu, and T. Lu, “A survey on fault diagnosis of rotating machinery based on machine learning,” *Measurement Science and Technology*, vol. 35, no. 10, p. 102001, 2024.
- [5] N. K. Hosseini, H. Toshani, S. Abdi, and S. Sharifzadeh, “Enhanced bearing fault detection in induction motors using projection-based svm,” *IEEE Transactions on Industry Applications*, 2025.
- [6] P. Arévalo, A. Cano, O. Fedoseienko, and F. Jurado, “A data-driven approach to microgrid fault detection and classification using taguchi-optimized cnns and wavelet transform,” *Applied Soft Computing*, vol. 170, p. 112667, 2025.
- [7] F. Lu, Q. Guo, Z. Dou, Y. Chen, Q. Wang, X. An, and H. Dou, “A novel simultaneous diagnosis method for igt open-circuit faults and current sensor faults of three-phase spwm inverter,” *IEEE Transactions on Power Electronics*, 2025.
- [8] Y. Wang, J. Shen, S. Yang, Q. Han, C. Zhao, P. Zhao, and X. Ren, “Knowledge and data dual-driven fault diagnosis in industrial scenarios: A survey,” *IEEE Internet of Things Journal*, 2024.
- [9] Y. Chen, F. Yu, L. Chen, G. Jin, and Q. Zhang, “Predictive modeling and multi-objective optimization of magnetic core loss with activation function flexibly selected kolmogorov-arnold networks,” *Energy*, p. 137730, 2025.
- [10] J. Yan, Y. Cheng, F. Zhang, M. Li, N. Zhou, B. Jin, H. Wang, H. Yang, and W. Zhang, “Research on multimodal techniques for arc detection in railway systems with limited data,” *Structural Health Monitoring*, p. 14759217251336797, 2025.
- [11] F. Feng, T. Zhu, B. Yang, Z. Zhang, S. Zhou, and S. Xiao, “Probabilistic fatigue life prediction in additive manufacturing materials with a physics-informed neural network framework,” *Expert Systems with Applications*, p. 127098, 2025.
- [12] H. Wang and Z. Liu, “Generalized koopman neural operator for data-driven modelling of electric railway pantograph-catenary system,” *IEEE Transactions on Transportation Electrification*, 2025.
- [13] H. Yang, Z. Liu, N. Ma, X. Wang, W. Liu, H. Wang, D. Zhan, and Z. Hu, “Csrn-mim: A self-supervised pre-training method for detecting catenary support components in electrified railways,” *IEEE Transactions on Transportation Electrification*, 2025.
- [14] Z. Zhang, B. Yang, F. Feng, S. Zhou, L. Yang, S. Xiao, and G. Yang, “Multiaxial fatigue model describing crack growth behavior and its application in welded structures of railway frames,” *International Journal of Fatigue*, p. 108831, 2025.
- [15] J. Yan, Y. Cheng, F. Zhang, N. Zhou, H. Wang, B. Jin, M. Wang, and W. Zhang, “Multi-modal imitation learning for arc detection in complex railway environments,” *IEEE Transactions on Instrumentation and Measurement*, 2025.
- [16] G. Han, J. Shen, Z. Wang, Y. Zhu, and Y. Xie, “A novel dual-domain adversarial method for vibration signal denoising in bearing fault diagnosis,” *IEEE Transactions on Instrumentation and Measurement*, 2025.
- [17] Y. Xiao, H. Shao, J. Wang, S. Yan, and B. Liu, “Bayesian variational transformer: A generalizable model for rotating machinery fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 207, p. 110936, 2024.
- [18] Z. Xu, X. Chen, J. Xu, and B. Zhao, “Multi-sensor signals augmented multimodal maml-1dcnn-rbeam deep transfer learning algorithm for wind turbine bearing fault diagnosis,” *Nondestructive Testing and Evaluation*, pp. 1–29, 2025.
- [19] J. Zhang, J. Chen, H. Deng, and W. Hu, “A novel framework based on adaptive multi-task learning for bearing fault diagnosis,” *Energy Reports*, vol. 9, pp. 522–531, 2023.
- [20] T. Man, V. Y. Osipov, N. Zhukova, A. Subbotin, and D. I. Ignatov, “Neural networks for intelligent multilevel control of artificial and natural objects based on data fusion: A survey,” *Information Fusion*, vol. 110, p. 102427, 2024.
- [21] J. Yan, Q. Wang, Y. Cheng, Z. Su, F. Zhang, M. Zhong, L. Liu, B. Jin, and W. Zhang, “Optimized single-image super-resolution reconstruction: A multimodal approach based on reversible guidance and cyclical knowledge distillation,” *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108496, 2024.

- [22] Y. Xu, K. Feng, X. Yan, R. Yan, Q. Ni, B. Sun, Z. Lei, Y. Zhang, and Z. Liu, “Cfcnn: A novel convolutional fusion framework for collaborative fault identification of rotating machinery,” *Information Fusion*, vol. 95, pp. 1–16, 2023.
- [23] Y. Xue, C. Wen, Z. Wang, W. Liu, and G. Chen, “A novel framework for motor bearing fault diagnosis based on multi-transformation domain and multi-source data,” *Knowledge-Based Systems*, vol. 283, p. 111205, 2024.
- [24] Y. Cheng, J. Yan, F. Zhang, M. Li, N. Zhou, C. Shi, B. Jin, and W. Zhang, “Surrogate modeling of pantograph-catenary system interactions,” *Mechanical Systems and Signal Processing*, vol. 224, p. 112134, 2025.
- [25] Z. Zhang, Y. Deng, and X. Liu, “A study on degradation prediction modeling method for mechanical rotating components based on multi-modal data and improved structured state-space sequence model,” *Journal of Intelligent Manufacturing*, pp. 1–21, 2025.
- [26] Z. Chen, H.-Z. Huang, Z. Deng, and J. Wu, “Shrinkage mamba relation network with out-of-distribution data augmentation for rotating machinery fault detection and localization under zero-faulty data,” *Mechanical Systems and Signal Processing*, vol. 224, p. 112145, 2025.
- [27] C. Wang, Y. Wang, F. Zheng, M. Li, and R. Xia, “Sstg: An interpretable spatio-temporal selective state-space model for multi-sensor data fusion in intelligent diagnosis,” *Knowledge-Based Systems*, vol. 316, p. 113278, 2025.
- [28] P. Ding, Y. Xu, and X.-M. Sun, “Multi-task learning for aero-engine bearing fault diagnosis with limited data,” *IEEE Transactions on Instrumentation and Measurement*, 2024.
- [29] J. Cui, P. Xie, X. Wang, J. Wang, Q. He, and G. Jiang, “M2fn: An end-to-end multi-task and multi-sensor fusion network for intelligent fault diagnosis,” *Measurement*, vol. 204, p. 112085, 2022.
- [30] H. Wang, Z. Liu, D. Peng, M. Yang, and Y. Qin, “Feature-level attention-guided multitask cnn for fault diagnosis and working conditions identification of rolling bearing,” *IEEE transactions on neural networks and learning systems*, vol. 33, no. 9, pp. 4757–4769, 2021.
- [31] Y. Cheng, N. Zhou, Z. Wang, B. Chen, and W. Zhang, “Cffsbd: A candidate fault frequencies-based blind deconvolution for rolling element bearings fault feature enhancement,” *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–12, 2023.
- [32] Y. Li, M. J. Zuo, Z. Chen, and J. Lin, “Railway bearing and cardan shaft fault diagnosis via an improved morphological filter,” *Structural Health Monitoring*, vol. 19, no. 5, pp. 1471–1486, 2020.
- [33] S. Wang, B. Chen, Y. Cheng, and X. Jiang, “Investigation on morphological filtering via enhanced adaptive time-varying structural element for bearing fault diagnosis,” *Measurement*, vol. 244, p. 116466, 2025.
- [34] B. Chen, D. Song, W. Zhang, Y. Cheng, and Z. Wang, “A performance enhanced time-varying morphological filtering method for bearing fault diagnosis,” *Measurement*, vol. 176, p. 109163, 2021.
- [35] L. Zhang, J. Xu, J. Yang, D. Yang, and D. Wang, “Multiscale morphology analysis and its application to fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 22, no. 3, pp. 597–610, 2008.
- [36] J. Guo, D. Zhen, H. Li, Z. Shi, F. Gu, and A. D. Ball, “Fault detection for planetary gearbox based on an enhanced average filter and modulation signal bispectrum analysis,” *ISA transactions*, vol. 101, pp. 408–420, 2020.
- [37] P. Borghesani, W. A. Smith, R. B. Randall, J. Antoni, M. El Badaoui, and Z. Peng, “Bearing signal models and their effect on bearing diagnostics,” *Mechanical Systems and Signal Processing*, vol. 174, p. 109077, 2022.
- [38] J. Antoni, G. Xin, and N. Hamzaoui, “Fast computation of the spectral correlation,” *Mechanical systems and signal processing*, vol. 92, pp. 248–277, 2017.
- [39] J. K. Alsalaet, “Fast averaged cyclic periodogram method to compute spectral correlation and coherence,” *ISA transactions*, vol. 129, pp. 609–630, 2022.
- [40] A. Mauricio and K. Gryllias, “Cyclostationary-based multiband envelope spectra extraction for bearing diagnostics: The combined improved envelope spectrum,” *Mechanical Systems and Signal Processing*, vol. 149, p. 107150, 2021.

- [41] S. Schmidt and K. C. Gryllias, "The anomalous and smoothed anomalous envelope spectra for rotating machine fault diagnosis," *Mechanical systems and signal processing*, vol. 158, p. 107770, 2021.
- [42] B. Zhang, Y. Miao, J. Lin, and H. Li, "Weighted envelope spectrum based on the spectral coherence for bearing diagnosis," *ISA transactions*, vol. 123, pp. 398–412, 2022.
- [43] L. Shen, Y. Hong, L. Li, H. Wang, and Q. Pu, "Physics-enhanced prediction model for shear bearing capacity of railway bridge piers based on flexural-shear coupling theory," *Engineering Structures*, vol. 343, p. 120993, 2025.
- [44] Y. Han, F. Zhang, Z. Li, Q. Wang, C. Li, P. Lai, T. Li, F. Teng, and Z. Jin, "Mt-convformer: A multitask bearing fault diagnosis method using a combination of cnn and transformer," *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1–16, 2025.
- [45] H. Chen, J. Wei, H. Huang, L. Wen, Y. Yuan, and J. Wu, "Novel imbalanced fault diagnosis method based on generative adversarial networks with balancing serial cnn and transformer (bctgan)," *Expert Systems with Applications*, vol. 258, p. 125171, 2024.
- [46] S. Gawde, S. Patil, S. Kumar, and K. Kotecha, "A scoping review on multi-fault diagnosis of industrial rotating machines using multi-sensor data fusion," *Artificial Intelligence Review*, vol. 56, no. 5, pp. 4711–4764, 2023.
- [47] Z. Liu, H. Wang, J. Liu, Y. Qin, and D. Peng, "Multitask learning based on lightweight 1dcnn for fault diagnosis of wheelset bearings," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.
- [48] X. Xiao, C. Li, H. He, J. Huang, and T. Yu, "Rotating machinery fault diagnosis method based on multi-level fusion framework of multi-sensor information," *Information Fusion*, vol. 113, p. 102621, 2025.
- [49] Y. He, H. Tang, Y. Ren, and A. Kumar, "A deep multi-signal fusion adversarial model based transfer learning and residual network for axial piston pump fault diagnosis," *Measurement*, vol. 192, p. 110889, 2022.
- [50] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 7, pp. 1693–1702, 2017.
- [51] H. Wang, S. Li, L. Song, and L. Cui, "A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals," *Computers in Industry*, vol. 105, pp. 182–190, 2019.
- [52] S. Hao, F.-X. Ge, Y. Li, and J. Jiang, "Multisensor bearing fault diagnosis based on one-dimensional convolutional long short-term memory networks," *Measurement*, vol. 159, p. 107802, 2020.
- [53] S. Guo, B. Zhang, T. Yang, D. Lyu, and W. Gao, "Multitask convolutional neural network with information fusion for bearing fault diagnosis and localization," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 9, pp. 8005–8015, 2019.
- [54] Z. Xie, J. Chen, Y. Feng, K. Zhang, and Z. Zhou, "End to end multi-task learning with attention for multi-objective fault diagnosis under small sample," *Journal of Manufacturing Systems*, vol. 62, pp. 301–316, 2022.
- [55] M. J. Hasan, M. M. Islam, and J.-M. Kim, "Bearing fault diagnosis using multidomain fusion-based vibration imaging and multitask learning," *Sensors*, vol. 22, no. 1, p. 56, 2021.
- [56] G. Niu, E. Liu, X. Wang, P. Ziehl, and B. Zhang, "Enhanced discriminate feature learning deep residual cnn for multitask bearing fault diagnosis with information fusion," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 1, pp. 762–770, 2022.
- [57] S. Chen, Z. Liu, X. He, D. Zou, and D. Zhou, "Multi-mode fault diagnosis datasets of gearbox under variable working conditions," *Data in Brief*, vol. 54, p. 110453, 2024.
- [58] S. Wan, T. Li, B. Fang, K. Yan, J. Hong, and X. Li, "Bearing fault diagnosis based on multisensor information coupling and attentional feature fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–12, 2023.
- [59] M. Ye, X. Yan, X. Hua, D. Jiang, L. Xiang, and N. Chen, "Mrcfn: A multi-sensor residual convolutional fusion network for intelligent fault diagnosis of bearings in noisy and small sample scenarios," *Expert Systems with Applications*, vol. 259, p. 125214, 2025.



Fan Zhang received the B.S. degree from Jincheng College of Sichuan University, Chengdu, China, in 2012, and the M.Sc. and Ph.D. degrees from the Department of Electronic and Computer Engineering of Brunel University London, London, U.K., in 2014 and 2019, respectively. He is currently a Lecturer with Southwest Jiaotong University, Chengdu, China. His research interests include deep learning, data fusion, fault diagnosis, and human-computer interaction.



Zhaoqi Li received the B.S. degree in mathematics from Southwest Jiaotong University, Chengdu, China, in 2023. He is currently pursuing the M.S. degree with Southwest Jiaotong University, Chengdu, China. His research interests include fault diagnosis.



Yao Cheng is engaged in the research of online monitoring and fault diagnosis for high-speed trains, focusing on temperature big data analysis. He received his Ph.D. in Engineering from Southwest Jiaotong University, Chengdu, China, in 2019, and is currently a researcher at the State Key Laboratory of Rail Transit Vehicle Systems, Chengdu, China.



Tianrui Li received his Ph.D. from Southwest Jiaotong University, Chengdu, China, in 2002. He is currently a Professor at the School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, China. His research interests include big data, cloud computing, data mining, and rough sets.



Jingke Yan is currently pursuing the Ph.D. degree at the State Key Laboratory of Rail Transit Vehicle Systems, Southwest Jiaotong University, Chengdu, China, since 2023. His research interests include deep learning.



Likai Dong received the B.S. degree from Shandong University, Jinan, China, in 2001 and the M.S. degree from the University of Jinan, Jinan, China, in 2008. His primary research interests include data mining and

intelligent information processing.



Yan Yang received the BS and MS degrees from Huazhong University of Science and Technology, Wuhan, China in 1984 and 1987, respectively. She received the PhD degree from Southwest Jiaotong University, Chengdu, China in 2007. She is currently a professor and vice dean at the School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, China. Her current research interests include multi-view learning, big data analysis, semi-supervised learning.



Yufei Han received the Bachelor's degree from the University of Jinan, Jinan, China in 2022 and the Master's degree from Southwest Jiaotong University, Chengdu, China in 2025. He is currently working at the Information System Integration Center of China Mobile Communications Group Shandong Company Limited, Jinan, China.



Cai Yi received her Ph.D. in vehicle engineering from Southwest Jiaotong University, Chengdu, China, in 2015. She was a Postdoctoral Research Fellow at the Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong, China, from 2017 to 2018. She is currently an Assistant Researcher at the State Key Laboratory of Traction Power, Southwest Jiaotong University, Chengdu, China, with research interests in vehicle component condition monitoring, fault detection, vibration signal processing, and machinery state prediction.



Weihua Zhang received his B.S. in mechanical automation, M.S. in engineering mechanics, and Ph.D. in vehicle operation engineering from Southwest Jiaotong University, Chengdu, China, in 1983, 1989, and 1996, respectively. He is currently a Professor at the State Key Laboratory of Rail Transit Vehicle Systems, Southwest Jiaotong University, Chengdu, China. His research interests include structural reliability, vehicle system dynamics and control, and measurement and test technology.